

# TWO PROBABILITY MODELS FOR SEQUENCES OF WET OR DRY DAYS

J. R. GREEN

University of Liverpool, Liverpool, Great Britain

## ABSTRACT

The goodness of fit of the Markov chain model to sequences of wet or dry days in data considered by Weiss is examined. Where appropriate an alternative probability model is also fitted to the data and the goodness of fit tested. The two models are found to give similar expected frequencies. An advantage of the alternative model is pointed out.

## 1. INTRODUCTION

Weiss [1] fitted a Markov chain model to the alternating sequences of wet days and dry days at a number of places, following the example of Gabriel and Neumann [2] who had done the same regarding the weather at Tel Aviv. Using data to which earlier writers had fitted different probability models, Weiss fitted the Markov model to data for Montsouris, San Francisco, Harpenden, Montreal, and Moncton (his table 1). He fitted it also to alternating sequences of days between and during stormy periods in four areas (his table 2), and to sequences of dry days at Kansas City, defined as having less than certain total rainfall amounts, four widely different amounts being considered (his table 3). Finally he fitted the model to data for each month of the year with data covering 50 years at Fort Worth (his table 4). He claimed that these fits were successful.

The present writer [3] found that the Tel Aviv data of Gabriel and Neumann were also well described by another probability model; in fact the new model fitted better in some respects. According to this model the probabilities of different run lengths for wet days or for dry days were derivable from the assumption that the alternating spells of continuous rain or continuous dryness formed an alternating renewal process in continuous time, each continuous spell having an exponential distribution. In the present paper this model is fitted to some of the sets of data in Weiss' paper. The model may be helpfully suggestive regarding the mechanism of rainfall occurrence at places where it seems to apply.

First of all each fit of the Markov model in Weiss' paper was tested by a  $\chi^2$  test. Some sets of data yielded significant results—that is, the discrepancies between the theoretical and observed frequencies were greater than could be safely attributed to random fluctuations supposing the Markov model to apply, but this model may yet give an approximate description of the situation good enough for some purposes. The alternating exponential model also gave a Markov process for runs of dry days.

In the three cases where there were data available for runs of dry days and runs of wet days and where the Markov model fitted the runs of dry days, *both* models fitted the data for runs of wet days. The simple Markov chain model used by Weiss considers persistence effects to be negligible beyond one day. To the extent that they are in fact not negligible and are taken into account by another probability model, we might well expect improvement in the fit. Nevertheless, the alternating exponential model, like the Markov chain model, only requires two parameters to be estimated from the data.

Thus, in those cases where the Markov process applied to both wet and dry days, so also did the alternating exponential process. As stated above, the latter is a Markov process for dry days, and it is also approximately so for wet days, especially for longer runs, as shown by Green [3]. The writer knows of no cases where one model applies and not the other. Where both models fit the data, the writer believes the alternating exponential model may be preferable because it suggests that the length of each spell of continuous dryness or rain has an independent exponential distribution with one parameter for dry spells and another for wet spells. This could be independently investigated for particular places where appropriate data are available, and if confirmed would be much more informative about the weather mechanism than the corresponding discrete model concerning days.

A graph is given showing when the two different models can be expected to give very similar results (fig. 1).

## 2. GOODNESS-OF-FIT TESTS

From the point of view of the  $\chi^2$  test the fit of the Markov model was unsatisfactory in the following parts of Weiss' paper: Montsouris, San Francisco, Harpenden, and Moncton of table 1, and Areas 2 and 3 of table 2. Parts C and D of table 3 hardly lend themselves to satisfactory  $\chi^2$  tests because of the large tail groupings and the need to quote the small individual expected frequencies more accurately than to the nearest integer. Further, in tables 1 to 4 it was necessary to adjust the

numbers in the tail groupings, because of accumulated rounding-off errors, to make the total observed and expected frequencies agree. Then the Markov model was found by the  $\chi^2$  test to give a satisfactory fit for the following parts of Weiss' paper: Montreal in table 1, Areas 1 and 4 in table 2, parts A and B in table 3, and all table 4.

Of those places where the Markov model fitted the weather data, only three had data available in Weiss' paper for wet days as well as dry days (or equivalently, stormy days as well as interval days in the case of table 2). These were Montreal of table 1 and Areas 1 and 4 of table 2. The alternating exponential model was found to give a good fit to the data for these three places. It should be explained that in fitting the model the maximum likelihood equations which should be solved to estimate two unknown parameters are rather intractable, but, using the fact that the Markov model also approximately applies, approximate maximum likelihood estimates can be obtained by equating the probability of a wet day following a given dry day and the probability of a wet day following a given wet day to the corresponding observed relative frequencies. These two probabilities are called  $p_0$  and  $p_1$ , respectively, by Weiss and shall be so called in this paper.

### 3. CONDITIONS FOR AGREEMENT BETWEEN THE TWO MODELS

Considering first the alternating exponential model, let  $P_n$  represent the probability of a wet day given that a dry day followed by  $n$  wet days immediately precede the day in question. It was shown by Green [3] that the  $P_n$ 's alternate about a limiting value, getting closer each time, as  $n$  increases. The largest  $P_n$  value is  $P_1$  and the smallest is  $P_2$ . The probability of a run of exactly  $n$  wet days, given such a run has just started, is  $P_1 P_2 \dots P_{n-1}(1-P_n)$ . If  $P_n$  had the same value for all  $n$ , this would be the same as the probability according to the Markov model, and in any case for longer runs, when the  $P_n$ 's for larger  $n$  are virtually constant, this model will approximately apply. The author has found in practice

TABLE 1.—Values of  $P_1$  and  $P_2$

$p_0 \backslash p_1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1 $P_1$	0.211	0.344	0.482	0.606	0.713	0.803	0.879	0.944
$P_2$	.153	.187	.266	.385	.519	.653	.779	.895
0.2 $P_1$		.311	.443	.575	.692	.790	.873	.942
$P_2$		.272	.325	.412	.529	.655	.779	.895
0.3 $P_1$			.411	.541	.667	.775	.865	.939
$P_2$			.381	.450	.544	.660	.780	.895
0.4 $P_1$				.511	.638	.757	.855	.935
$P_2$				.487	.567	.667	.782	.895
0.5 $P_1$					.611	.734	.844	.931
$P_2$					.591	.680	.784	.895
0.6 $P_1$						.711	.828	.926
$P_2$						.694	.789	.896
0.7 $P_1$							.810	.918
$P_2$							.796	.897
0.8 $P_1$								.908
$P_2$								.898

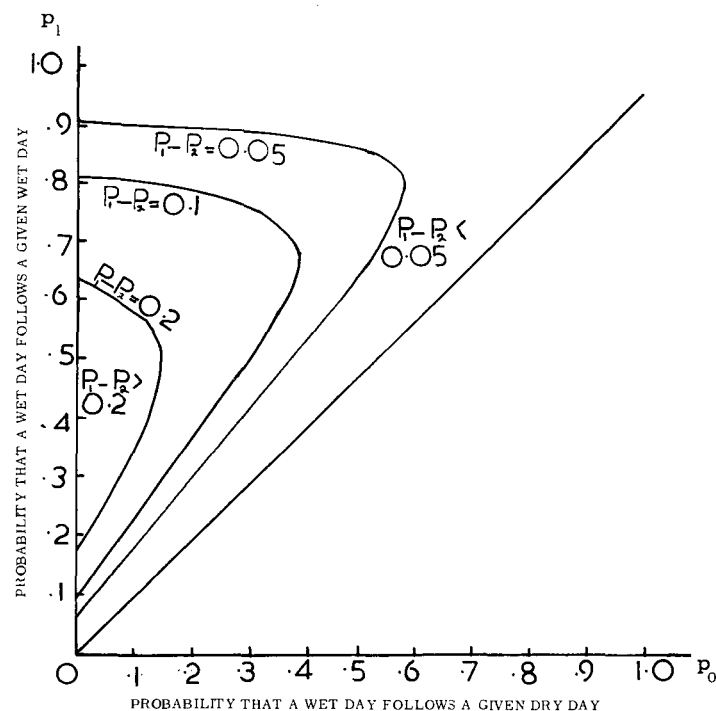


FIGURE 1.—Contours of  $P_1 - P_2$ , low values of which indicate good agreement between the two probability models of this paper. (Empirical values of  $p_1$  and  $p_0$  are entered. The points should lie above the diagonal; i.e.,  $p_1 > p_0$ .)

that the  $P_n$ 's quickly become almost constant with increasing  $n$ , so that the two models will often be in close agreement. Apart from that general observation, it is evident that the closeness of  $P_1$  and  $P_2$  is a good indication of the closeness of agreement between the two models. Figure 1 shows some contours of  $P_1 - P_2$  for different values of  $p_0$  and  $p_1$ . Using this figure, knowing the values of  $p_0$  and  $p_1$  estimated from observed data for a certain place, one can obtain an indication of whether or not the two models are likely to be in close agreement if fitted to the data for that place. Table 1 shows the actual values of  $P_1$  and  $P_2$  for some values of  $p_0$  and  $p_1$ . In the alternating exponential model  $p_1$  is necessarily greater than  $p_0$ .

### REFERENCES

1. L. L. Weiss, "Sequences of Wet or Dry Days Described by a Markov Chain Probability Model," *Monthly Weather Review*, vol. 92, No. 4, Apr. 1964, pp. 169-176.
2. K. R. Gabriel and J. Neumann, "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," *Quarterly Journal of the Royal Meteorological Society*, vol. 88, No. 375, Jan. 1962, pp. 90-95.
3. J. R. Green, "A Model for Rainfall Occurrence," *Journal of the Royal Statistical Society, Series B*, vol. 26, No. 2, 1964, pp. 345-353.